Exhibit A

Gene, 25 (1983) 263–269
Elsevier

GENE 876

A simple and very efficient method for generating cDNA libraries

(Recombinant DNA; oligo(dT)-primed first strand synthesis; RNase H and DNA polymerase I mediated second strand synthesis; full length clones; globin; preproenkephalin)

Ueli Gubler and Beth J. Hoffman

Department of Molecular Genetics, Roche Research Center, Hoffmann-La Roche Inc., Nutley, NJ 07110 (U.S.A.) Tel. (201) 235-5554

(Received July 6th, 1983)
(Accepted July 14th, 1983)

SUMMARY

A simple method for generating cDNA libraries from submicrogram quantities of mRNA is described. It combines classical first-strand synthesis with the novel RNase H-DNA polymerase I-mediated second-strand synthesis [Okayama, H., and Berg, P., Mol. Cell. Biol. 2 (1982) 161–170]. Neither the elaborate vector-primer system nor the classical hairpin loop cleavage by S1 nuclease are used. cDNA thus made can be tailed and cloned without further purification or sizing. Cloning efficiencies can be as high as 10^6 recombinants generated per μg mRNA, a considerable improvement over earlier methods. Using the fully sequenced 1300 nucleotide-long bovine preproenkephalin mRNA, we have established by sequencing that the method yields faithful full-length transcripts. This procedure considerably simplifies the establishment of cDNA libraries and thus the cloning of low-abundance mRNAs.

INTRODUCTION

Since the first reports on the molecular cloning of cDNAs (Rougeon and Mach, 1976; Efstratiadis et al., 1976), this technology has become more and more refined. It has evolved into a powerful and versatile tool in the molecular analysis of both eukaryotic and prokaryotic genes. Recent modifications have made it possible to clone and express mRNAs which encode proteins that would otherwise be difficult to purify in large amounts. The mRNAs for these polypeptides are often low in abundance. Therefore, to statistically have a chance of finding the desired sequence, a technique which generates large numbers of clones from small amounts of template mRNA is essential. To develop a more efficient cDNA cloning system, efforts have focused on eliminating the S1-nuclease-mediated cleavage of the hairpin loop (Land et al., 1981; Okayama and Berg, 1982), since this step introduces considerable loss resulting in rather low numbers of clones generated per μg of mRNA and also often removes important 5'-terminal sequences from the cloned mRNAs. Okayama and Berg (1982) describe an elegant method which circumvents the hairpin loop cleavage by using a specifically designed plasmid-vector primer. Briefly

Abbreviations: AMV, avian myeloblastosis virus; bp, base pairs; BSA, bovine serum albumin; β-NAD, β-nicotinamide adenine dinucleotide; cDNA, DNA complementary to mRNA; DTT, dithiothreitol; kb, kilobase pairs; LB, Luria broth; SDS, sodium dodecyl sulfate; TCA, trichloroacetic acid.

outlined, their cloning system is set up as follows: (1) the mRNA is transcribed into cDNA using T-tailed cloning vector as a primer; (2) annealing and ligation of a linker fragment to the recombinant hybrid molecules thus produced ensures their circularization; (3) the second cDNA strand is then synthesized by replacing the RNA in the hybrids with DNA by using a combination of RNase H, DNA polymerase I and DNA ligase before the molecules are introduced into competent *Escherichia coli* cells.

We now report a simple modification of the Okayama and Berg (1982) protocol. Our method combines (1) classical oligo(dT)-primed first-strand synthesis with (2) the novel RNase H-DNA polymerase I-mediated second strand synthesis. Neither the elaborate vector-primer system nor the classical hairpin-loop cleavage by S1 nuclease are used. Thus, cDNA can be synthesized in two simple steps and cloned without further sizing. In our hands, this procedure yields cloning efficiencies of up to 10^6 clones/μg mRNA.

MATERIALS AND METHODS

(a) Enzymes

Enzymes and chemicals were purchased from the following suppliers: AMV reverse transcriptase, Life Sciences Inc.; DNA polymerase I, Boehringer Mannheim Biochemicals (5000 units/mg) or New England Biolabs; *E. coli* DNA ligase and restriction enzymes, New England Biolabs; RNase H and rabbit reticulocyte poly(A)$^+$ RNA, BRL; terminal deoxynucleotidyl transferase, oligo(dT_{12-18}) and oligo(dT)-cellulose (type 7), PL-Biochemicals; bovine serum albumin (fraction V), Pentex®, Miles.

(b) DNA and RNA manipulations

Total RNA is prepared by extraction of tissues or cultured cells with either proteinase K-SDS (Hall et al., 1979) or guanidine thiocyanate (Chirgwin et al., 1979). Poly(A)$^+$ RNA is selected on oligo(dT)-cellulose. Transformation of *E. coli* RR1 with chimeric molecules is as described by Peacock et al. (1981). Mini-lysates are prepared according to Birnboim and Doly (1979). Alkaline agarose gels are run according to a published procedure of McDonell et al. (1977). The chemical sequencing method is used as described by Maxam and Gilbert (1980).

(c) cDNA synthesis

Synthesis of first-strand cDNA is carried out in a reaction volume of 20–40 μl containing 50 mM Tris · HCl, pH 8.3, 10 mM $MgCl_2$, 10 mM DTT, 4 mM Na · pyrophosphate, 1.25 mM dGTP, 1.25 mM dATP, 1.25 mM TTP, 0.5 mM dCTP, 15–20 μCi of [α-^{32}P]dCTP ($\simeq$ 3000 Ci/mmol), 100 μg/ml of oligo(dT_{12-18}), 150 μg/ml poly(A)$^+$-RNA, 3000 units reverse transcriptase/ml for 30 min at 43°C. The reaction is stopped by adding EDTA to 20 mM. The products are extracted with phenol and precipitated with ethanol out of 2 M NH_4 · acetate as described by Okayama and Berg (1982). The amounts of first strands synthesized are estimated by assaying TCA-insoluble radioactivity. For second strand synthesis, up to 500 ng of single-stranded cDNA (i.e., 1 μg of hybrid) can be processed in 100 μl of 20 mM Tris · HCl, pH 7.5, 5 mM $MgCl_2$, 10 mM $(NH_4)_2SO_4$, 100 mM KCl, 0.15 mM β-NAD, 50 μg/ml BSA, 40 μM dNTPs, 8.5 units/ml of *E. coli* RNase H, 230 units/ml DNA polymerase I, 10 units/ml *E. coli* DNA ligase. Incubations are sequentially 60 min at 12°C and 60 min at 22°C. EDTA is added to 20 mM to stop the reaction. The products are extracted with phenol twice and then precipitated out of 2 M NH_4 · acetate twice as described above. This procedure yields double-stranded cDNA that can be tailed and cloned without further sizing. Analytical second-strand synthesis with reverse transcriptase is carried out as described by Myers et al. (1980).

(d) Cloning

A cloning vector with a low transformation background is conveniently prepared as follows: 200 μg of pBR322 are cut to completion with *Eco*RV. The DNA is purified by phenol extraction and is then sedimented through a sucrose gradient (Norgard et al., 1980). Gradient fractions are assayed for A_{260} and for transformation background. Fractions containing DNA with low transformation efficiency are pooled and concentrated by ethanol precipitation. It is our experience that linearization of the

plasmid without subsequent gradient purification yields a cloning vector whose transformation background is still on the order of about 10 colonies/ng. Gradient-purified vector, on the other hand, yields about 1 colony/ng or less (supercoil efficiency ≈ 2000–3000 colonies/ng). Tailing of this vector is carried out under the following conditions: 200 mM K · cacodylate, pH 6.9, 1 mM $CoCl_2$, 0.9 mM dCTP, 5 mg/ml BSA, 1 mg/ml *Eco*RV-cut pBR322, 1000 units terminal transferase/ml for 50 min at 22 °C. The enzyme is then heat-inactivated for 10 min at 65 °C. The vector is used as such without further purification. Tailing of the cDNA with dGTP is carried out under similar conditions. We usually use 100 ng of cDNA in 40 µl, with 30 units of terminal transferase for 60 min at 37 °C, followed by heat inactivation. Under these conditions, 20–30 nucleotides are added to each 3'-end, a length that is required for maximum transformation efficiency (Peacock et al., 1981). To find the maximum number of clones obtainable per weight input of cDNA, we usually titrate several annealing ratios (i.e. weight ratios vector:cDNA). For cDNA derived from cellular poly(A)$^+$ RNA and thus heterogeneous in size, ratios of 10:1, 20:1, 50:1, 100:1, have proven to be useful. Annealing is conveniently done in 50 µl of 10 mM Tris · HCl, pH 7.5, 1 mM EDTA, 150 mM NaCl at total DNA concentrations of 0.5 µg/ml or less (Peacock et al., 1981) for 60–90 min at 58 °C. Transformation is carried out by adding 100 µl of $CaCl_2$-competent cells and subsequent plating onto LB + 100 µg/ml ampicillin.

RESULTS AND DISCUSSION

(a) Globin cDNA

We used rabbit reticulocyte poly(A)$^+$ RNA to test the validity of our approach. RNA was reverse-transcribed without incorporating any labelled triphosphates. Yields for first-strand synthesis were estimated by running an identical parallel reaction which included [α-^{32}P]dCTP. The RNA in the unlabelled hybrids was subsequently replaced with DNA using a variety of different protocols. Incorporation of label into the second strand was monitored by analyzing all the samples after alkaline hydrolysis on alkaline gels in order to determine whether the RNA had been fully replaced by DNA. Fig. 1 shows that incorporation of label into a full length globin DNA was clearly dependent on the addition of both RNase H and DNA polymerase I (lanes 4 and 6). DNA polymerase I is used under conditions of nick translation to ensure complete replacement of RNA segments by DNA (Kornberg, 1980). The size of the second strands thus synthesized was identical to the size of the first strands generated with reverse transcriptase (lane 1). The RNase H activity intrinsic to the reverse transcriptase (Myers et al., 1980) is not sufficient to generate enough RNA primers for the DNA polymerase I (lanes 3, 7). This RNase H activity varies from batch to batch of AMV-reverse-transcriptase (G.E. Houts, personal communication). In order to get consistent results, we have found it necessary to add exogeneous RNase H. As demonstrated in lane 6, addition of DNA ligase is unnecessary. Although all of the data presented below have been obtained with DNA ligase in the second strand synthesis reaction, we have since

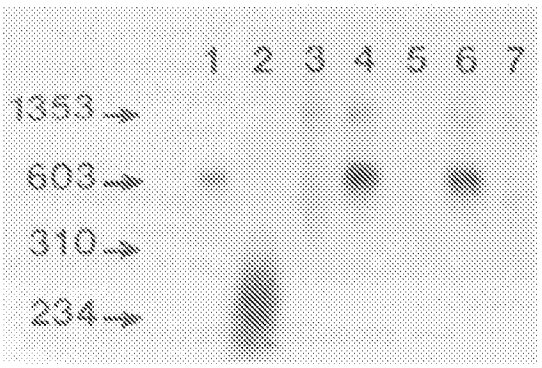

Fig. 1. Analysis of second-strand synthesis using different combinations of RNase H, DNA polymerase I and DNA ligase. The conversion of unlabelled globin mRNA cDNA hybrid into double-stranded cDNA was followed by incorporation of [α-^{32}P]dCTP into the second strands. All the samples were incubated for 30 min in 0.3 N NaOH at 46 °C prior to electrophoresis on a 1% alkaline agarose gel. The gel was dried and autoradiographed. Markers were fragments of φX174 DNA cut with *Hae*III, run in a parallel lane and stained with ethidium bromide (bands not shown; see arrows and numerals which specify bp). Lane 1: control, labelled first strands; lane 2: second-strand synthesis with reverse transcriptase only; lane 3: second-strand synthesis with DNA polymerase I only; lane 4: second-strand synthesis with RNase H, DNA polymerase I, DNA ligase; lane 5: no DNA polymerase I; lane 6: no DNA ligase; lane 7: no RNase H.

found that cDNA libraries can be successfully established without adding this enzyme (unpublished results). Reverse-transcriptase alone can be used for second strand synthesis, but the products generated are not full length (lane 2). This finding agrees with published reports (Myers et al., 1980; Boone and Skalka, 1981). We have not attempted to clone such molecules.

Analytical S1 digestion of globin cDNA that was rendered double-stranded in the standard reaction indicates that about 70% of the incorporated radioactivity was S1-resistant. For simplicity, we normally do not assay S1-resistance but assume full conversion from hybrid to double-stranded DNA. Weight inputs of cDNA are then calculated based on this assumption. Some of the products seen in Fig. 1 are clearly twice the original mRNA size of about 600 nucleotides (cf. lanes 3, 4, 6, 7). In these molecules, the hairpin loop on the first strand may have been used as a priming structure for the second-strand synthesis. Of course, such molecules can only be detected when using a homogeneous template like globin mRNA. The total amount of these oversized molecules varied between experiments. It was usually on the order of a few percent. We have not been able to account for these variations. Since no S1-nuclease treatment will follow, such molecules will not be tailed and hence will not be cloned.

After second-strand synthesis, the globin cDNA was tailed with dGTP and annealed to *Eco*RV-cut, C-tailed pBR322. Cloning efficiencies of about 600 000 clones generated per μg of mRNA were obtained in these experiments. Expressing the efficiency in this way has the advantage that the actual extent of conversion from hybrid to double-stranded DNA becomes irrelevant. Analysis of ten randomly picked globin clones demonstrated the following (not shown): (1) five out of ten clones were full length as judged by restriction-enzyme analysis; (2) two of the ten clones had a *Bam*HI site in their insert, which is characteristic of the rabbit β-globin cDNA (Efstratiadis et al., 1977); (3) sequence analysis of one full-length α-clone and one full-length β-clone showed that these recombinants were only missing 14 and 15 nucleotides, respectively, of the published 5′-untranslated region adjacent to the cap structure of the mRNAs.

(b) Properties of the *Eco*RV site on pBR322

The pBR322 vector used in these studies has a novel feature. Sequencing of clones generated with this vector has shown that the *Eco*RV site closer to the *Eco*RI site in the plasmid has been converted to a *Bam*HI site. Re-examination of the *Eco*RV cleavage site has shown that *Eco*RV generates blunt ends rather than 3′-overhanging ends (I. Schildkraut, New England Biolabs, personal communication). By looking at the structure and flanking sequences of the *Eco*RV site in pBR322, it becomes apparent that the C-tailing of the blunt ends will, indeed, create a *Bam*HI site (Fig. 2). cDNA inserts cloned into this vector are thus conveniently sized by cutting the recombinants with *Bam*HI alone. The

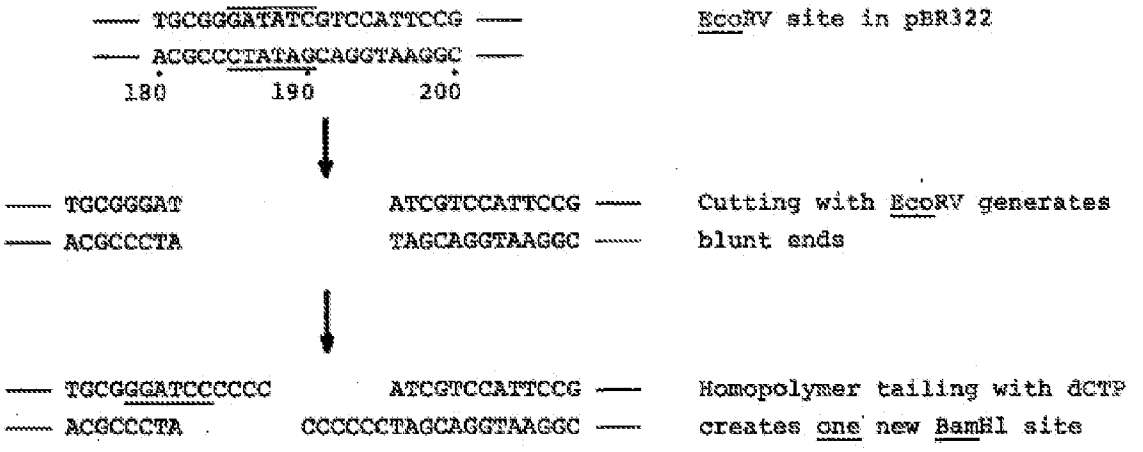

Fig. 2. The generation of the new *Bam*HI site in the pBR322-*Eco*RV cloning vector used in our studies. Restriction analysis and sequencing of several recombinant clones has unambiguously demonstrated that the new *Bam*HI site is present (not shown).

insert fragment will contain 188 bp of DNA from pBR322, stretching from the *Eco*RV site (pos 187) to the *Bam*HI site (pos 375).

Our original incentive for using *Eco*RV to prepare a cloning vector was, of course, the assumption that the tailing would recreate the restriction site. Even though that strategy failed, the newly created *Bam*HI site is very useful for restriction site mapping. Pure insert fragment can be conveniently generated by using one additional site in the insert.

(c) Preproenkephalin cDNA

Next, as a more stringent test of our cloning procedure, we generated cDNA clones with a longer and much less abundant mRNA, bovine preproenkephalin mRNA. This mRNA is about 1300 nucleotides long, constitutes approx. 0.1% of the total poly(A)$^+$ RNA in bovine adrenal medulla and has been completely sequenced using conventional cDNA cloning techniques (Gubler et al., 1982; Noda et al., 1982). Total poly(A)$^+$ RNA from bovine adrenal medulla was enriched for the 1–3 kb size class on a sucrose gradient. A cDNA library was established from this RNA by the outlined procedure. The small scale titration of several vector:cDNA ratios is illustrated in Table I. The final library was established by scaling up experimental point No. 2 100-fold. Using a final amount of 30 ng of double-stranded, G-tailed cDNA (derived from 200 ng of mRNA) and 1.5 μg of pBR322 vector, we were able to generate about 200 000 clones. This gives an efficiency of about 10^6 clones/μg mRNA. The vector-derived background was on the order of 0.2‰.

Screening of this new library with a proenkephalin cDNA from earlier work (Gubler et al., 1982) allowed us to isolate several new proenkephalin clones. Their inserts were analyzed by digestion with *Bam*HI. Fig. 3 shows that eight of those twelve clones had an insert of 1 kb or longer. Since the published length of preproenkephalin mRNA is 1223 nucleotides [excluding the poly(A)-tail], full length clones should contain about 1500 bp, resulting from a *Bam*HI insert of 1223 bp + 190 bp (the pBR-derived fragment) + approx. 100 bp [the cloned portion of the poly(A)-tail plus the GC-tails]. Clones No. 921 (lane 10) and No. 1021 (lane 12) fulfilled this criterion. They were further studied by chemical sequencing (Fig. 4). Since the initiator AUG codon in preproenkephalin mRNA is part of an *Nco*I site (Noda et al., 1982), the recombinants were cut with *Nco*I, labelled by either filling in with Klenow fragment of *E. coli* DNA polymerase I, or by kinasing with polynucleotide kinase. Subsequent cutting with *Bam*HI generated singly end-labelled fragments. These fragments were then sequenced. The results (Fig. 4) show that one of our clones, No. 1021, is identical in size to the published preproenkephalin cDNA (Noda et al., 1982). Primer extension experiments had indicated that the mRNA is about 65 nucleotides longer than the cDNA insert contained in No. 1021 (Noda et al., 1982). Since it is approximately 50 nucleotides longer than recombinant No. 1021, No. 921 has thus a very good chance of

TABLE I

Effect of weight ratios (A/B) of (A) C-tailed pBR322 vector to (B) adrenal medullary cDNA on the cloning efficiency

Annealing and transformation were done as outlined in MATERIALS AND METHODS, section d. Using supercoiled pBR322, the transformation efficiency was 2×10^6 colonies/μg. The vector-derived background was about 1 colony/ng (Exp. No. 1).

Exp. No.	A ng vector	B ng cDNA	A/B approx. ratio	Number of colonies obtained	Efficiency (colonies/ng cDNA)
1	15	–	–	18	–
2	15	0.3	50	337	1123
3	15	0.6	25	230	383
4	15	1.2	12.5	605	504
5	15	2.4	6	448	187
6	15	4.8	3	684	143
7	15	9.6	1.5	157	16

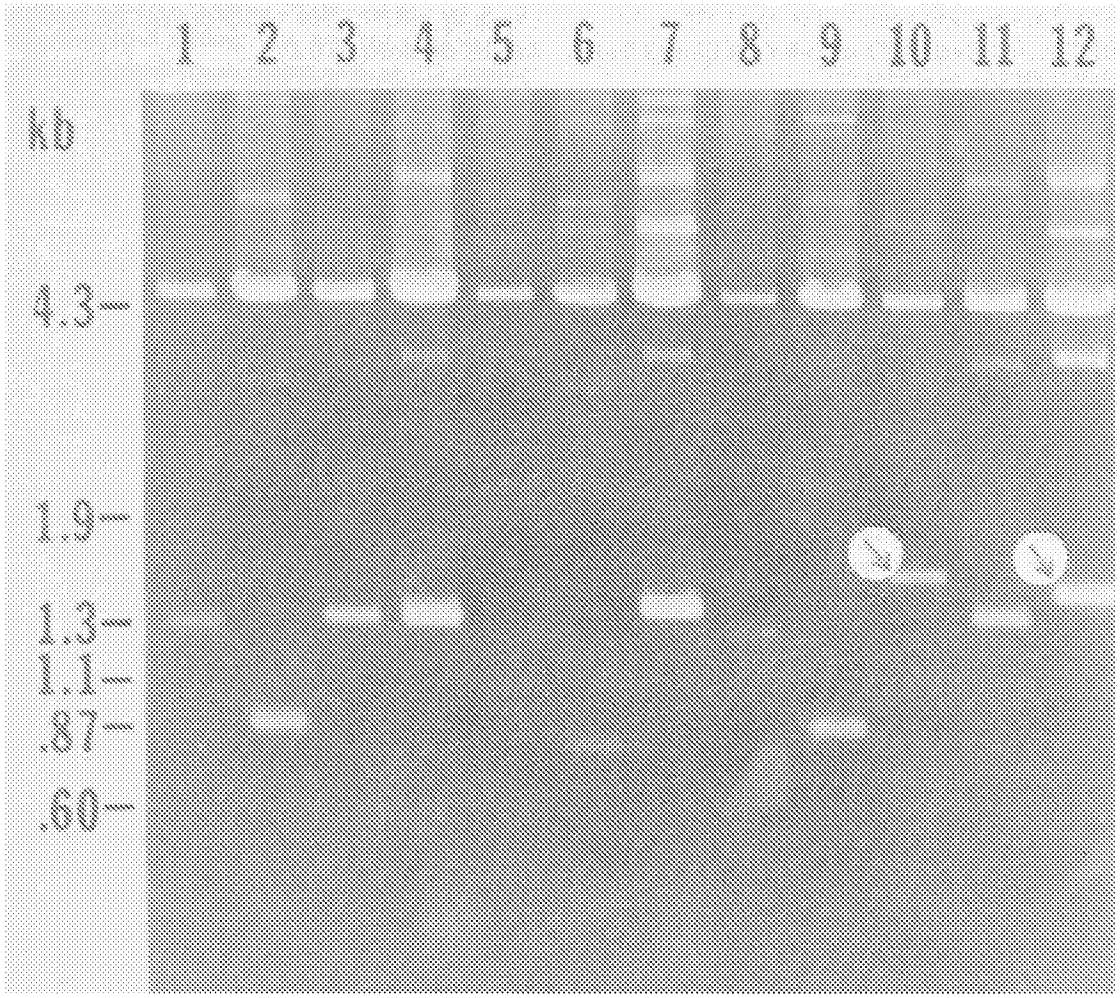

Fig. 3. Analysis of preproenkephalin clones by restriction mapping. Twelve recombinants were analyzed by digestion with *Bam*HI. A full-length preproenkephalin-cDNA insert is expected to be ⩾ 1500 bp long. Clones No. 921 (lane 10) and No. 1021 (lane 12) fulfill that criterion. Their inserts are marked with an arrow in a white circle in the figure. Markers were λ DNA digested with *Hin*dIII and φX174 DNA digested with *Hae*III.

```
        10         20         30         40         50         60         70         80         90
AGGCGGCGAG GACCGAGACC GCTCTGCGCT GCAGCCGCGG CGACGCCGAG GACCGCGAGA GTGAGGCCCG CCCGCTTTCC TGGCTCTCCC

                                                   AG GACCGCCAGA GTGAGGCCCG CCCGCTTTCC TGGCTCTCCC


      100        110        120        130        140        150
CTCGCCGAGA GTCGCCCCGG ACCCGGTTTC CACGACCGAC CTGCGTGCCC CGAACAGCGG CAA                          .....A

CTCGCCGAGA GTCGCCCCGG ACCGGGTTTC CACGACCGAC CTGCGTGCCC CGAACAGCGG CAACCCCATG GCG CGG TTC  .....B
                         *                                                  MET ALA ARG PHE
```

Fig. 4. Partial sequencing of full-length clones Nos. 921 and 1021 (line A) and comparison to the published sequence of the 5'-terminal region of preproenkephalin mRNA (line B). Both strands were sequenced for each clone. The two sequences are identical except for a single base change. In position 114 (*), there is a C in clone No. 921 and a G in clone No. 1021. Nucleotides 1–48 represent new sequence information derived from clone 921. Clone 1021 and the published preproenkephalin sequence are identical in length. The initiator codon within the *Nco*I site, which was end-labelled for sequencing, is underlined (see RESULTS AND DISCUSSION, section c). Using the same labelled site for both clones, 153 nucleotides were also sequenced into the coding region (to amino acid residue No. 51). This demonstrated complete agreement between the published sequence and our clones (not shown).

representing a complete cDNA transcript of preproenkephalin mRNA. These data demonstrate that our method is able to yield complete cDNA copies of considerable size that are faithful transcripts of their respective mRNAs.

(d) Efficiency of cloning

So far, we have used this procedure to generate about half a dozen libraries from different mRNA sources. We have consistently obtained efficiencies of 300 000 clones/μg mRNA or greater. In our hands, this represents an improvement of about 10- to 50-fold over what we have obtained with the S1 method. Our experience is that size enrichment of the poly(A)$^+$ RNA over a sucrose gradient helps to improve yields and transcript lengths considerably. The longest clone obtained to date has an insert of 2.4 kb and was made from mRNA that was enriched for a size class of 1.5 to 3 kb. Unfractionated poly(A)$^+$ RNA routinely gives rise to cDNA inserts that are up to 1.2 kb in size.

This procedure considerably simplifies previous methods and thus facilitates the cloning of rare mRNAs.

ACKNOWLEDGEMENTS

We would like to thank Dr. Fred Esch for performing the sequence analysis of the globin clones.

REFERENCES

Birnboim, H.C. and Doly, J.: A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7 (1979) 1513–1523.

Boone, L.R. and Skalka, A.-M.: Viral DNA synthesized in vitro by avian retrovirus particles permeabilized with melittin, II. Evidence for a strand displacement mechanism in plus-strand synthesis. J. Virol. 37 (1981) 117–126.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J.: Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18 (1979) 5294–5299.

Efstratiadis, A., Kafatos, F.C., Maxam, A.M. and Maniatis, T.: Enzymatic in vitro synthesis of globin genes. Cell 7 (1976) 279–288.

Efstratiadis, A., Kafatos, F.C. and Maniatis, T.: The primary structure of rabbit β-globin mRNA as determined from cloned cDNA. Cell 10 (1977) 571–585.

Gubler, U., Seeburg, P., Hoffman, B.J., Gage, L.P. and Udenfriend, S.: Molecular cloning establishes proenkephalin as precursor of enkephalin-containing polypeptides. Nature 295 (1982) 206–208.

Hall, L., Craig, R.K. and Campbell, P.N.: mRNA species directing synthesis of milk proteins in normal and tumor tissue from human mammary gland. Nature 277 (1979) 54–56.

Kornberg, A.: DNA Replication. Freeman, San Francisco, CA, 1980, pp. 101–166.

Land, H., Grez, M., Hauser, H., Lindenmaier, W. and Schütz, G.: 5'-Terminal sequences of eukaryotic mRNA can be cloned with high efficiency. Nucl. Acids Res. 9 (1981) 2251–2266.

Maxam, A. and Gilbert, W.: Sequencing end-labelled DNA with base-specific chemical cleavages, in Grossman, L. and Moldave, K. (Eds.), Methods in Enzymology, Vol. 65. Academic Press, New York, 1980, pp. 499–560.

McDonell, M.W., Simon, M.N. and Studier, F.W.: Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol. 110 (1977) 119–146.

Myers, J.C., Dobkin, C. and Spiegelman, S.: RNA primer used in synthesis of anticomplementary DNA by reverse transcriptase of avian myeloblastosis virus. Proc. Natl. Acad. Sci. USA 77 (1980) 1316–1320.

Noda, M., Furutani, Y., Takahashi, H., Toyosato, M., Hirosi, T., Inayama, S., Nakanishi, S. and Numa, S.: Cloning and sequence analysis of cDNA for bovine adrenal preproenkephalin. Nature 295 (1982) 202–206.

Norgard, M.V., Tocci, M.J. and Monahan, J.J.: On the cloning of eukaryotic total poly(A)-RNA populations in *Escherichia coli*. J. Biol. Chem. 255 (1980) 7665–7672.

Okayama, H. and Berg, P.: High-efficiency cloning of full-length cDNA. Mol. Cell. Biol. 2 (1982) 161–170.

Peacock, S.L., McIver, C.M. and Monahan, J.J.: Transformation of *E. coli* using homopolymer-linked plasmid chimeras. Biochim. Biophys. Acta 655 (1981) 243–250.

Rougeon, F. and Mach, B.: Stepwise biosynthesis in vitro of globin genes from globin mRNA by DNA polymerase of avian myeloblastosis virus. Proc. Natl. Acad. Sci. USA 73 (1976) 3418–3422.

Communicated by A.M. Skalka.